Spring 2000

Partial Amplification and Cloning of a Beta Carbonic Anhydrase in Arabidopsis thaliana

Zsofia Banhegyi

Spring 2000
Honors Research Thesis

Partial Amplification and Cloning of a Beta Carbonic Anhydrase in *Arabidopsis thaliana*

Written by: Zsofia Banhegyi
Research director: Dr. Sue G. Bartlett

1. Abstract

The goal of this research project was dual. The first goal was to find the full length Open Reading Frame (ORF) of a CA-like gene in *Arabidopsis thaliana*. The second goal was to amplify the promoter sequence and to express the promoter with Gus fusion in plants to learn about the spatial pattern of expression of the CA-like gene. Sequence alignment technique was used to find a proposed sequence for the CA-like gene in *Arabidopsis*.

We searched for the sequence in a cDNA library and isolated mRNA. Although Polymerase Chain Reaction and gel electrophoresis seemed to indicate that the CA-like gene was found in the cDNA library, sequencing results showed no homology to the proposed sequence. When RNA library was used the 554 base pair sequence amplified by Reverse Transcriptase PCR (RT-PCR) matched the proposed sequence. In addition we were able to learn about the splicing of this DNA segment. The sequence amplified did not contain introns because the template was RNA. When the bases were aligned to the proposed sequence we were able to deduct which segments were spliced out. As of right now we have the 3' end of the ORF and we are working to find the 5' end of the CA-like gene.

The search for the promoter sequence is still in process. Several PCR techniques were used to amplify the upstream sequence. When we get the correct sequence we will be able to tell whether the proteins encoded by the sequence are expressed in the chloroplast or the cytoplasm. We will also be able to tell what part of the plant this protein is produced.

2. Introduction

Carbonic anhydrase (CA) is an enzyme found in almost every organism. It catalyzes the interconversion of CO2 and HCO3-. It is an oligomeric metalloenzyme. It requires zinc for activity, which is bound to the catalytic site by three histidine residues (1). According to the sequences of the genes encoding this enzyme this protein can be broken down to three families: α, β, and γ. Earlier it was thought that the α family only existed in animals, the β family in chloroplasts and eubacteria, and γ in archaebacteria (2). Today we know that all three families can be represented in a single organism. For example in *Arabidopsis thaliana*, which is the plant we studied, there are at least 6 different α, 5 β, an 3 γ CAs. The gene chosen for my thesis research shares the characteristics of a β CA; sequence alignments indicated that important cysteine and histidine residues line up with that of other β CAs (3)(See page 8). The gene is located on chromosome #4.

B CAs have been found in the stroma of higher plant chloroplasts, C3 plants, and in eubacteria (4). In photosynthetic organisms β CAs can have four

different roles (5). The first role is to convert HCO3- into CO2 to supply the enzymes that require carbon-dioxide as a substrate. The second function is to convert CO2 to HCO3- to provide substrate for enzymes such as phoshoenolpyruvate carboxylase. The third activity of this enzyme is to promote the diffusion of CO2 across the plasma membrane and the chloroplast membrane. Finally CA makes the CO2 active transport through the plasma membrane possible by converting it to HCO3-(5).

The reason this particular gene was chosen for the research project is that Dr. Bartlett's laboratory has a particular interest in CA's. In fact her lab was the first one to identify CA in spinach (6). This discovery had a major importance because CA in spinach and in other plants has almost no sequence homology to CA's in animals. Interestingly however, crystal structures showed that the active sites of the mammalian and spinach CAs were virtually superinposable (1). A cysteine residue at the active site was deemed indispensable for CA activity by Dr. Barlett's lab (1).

A proposed sequence for a CA-like gene in *Arabidopsis* was found in the GenBank on the World Wide Web (7). This sequence was proposed by the ESSA group

it was given the identification number T16L1.70. The CA-like gene was about 940 base pairs long (see page 7.) When it was translated it seemed to be missing the cysteine residue that was thought to be necessary for activity by Dr. Bartlett's laboratory (see page 8). Therefor when the project was started we were not sure if this sequence was in fact coding for a CA or if the cysteine residue was dispensable for its activity.

In order to find this CA-like sequence in *Arabidopsis* we used a cDNA library and isolated mRNA. The cDNA library contains DNA copies of mRNA (expressed genes) cloned into a phage vector called lamdaGEM2. The technique to find the sequence in question is called "Nested PCR." This means that first the random segment that includes the multiple cloning site but excludes the "lambda-arms" of the phage is amplified by non specific primers. Then primers are designed to hybridize with the 5' end and the 3' end of the desired sequence. Originally we did not know where the CA-like gene began. Therefore two 5' primers were designed. With this arrangement one set of primers flanked a 533 base pair region (AtCA-like-1 at the 5' end and AtCA-like-3 at the 3' end). The other set of primers flanked a 653 base pair region including the 533 base

pair region at the 3′ end (AtCA-like2 at the 5′ end and AtCA-like-3 at the 3′ end.) This way the longer sequence included the cysteine residue in question while the shorter segment excluded it.

Our efforts seemed to fail when the cDNA library was used. However when the isolated mRNA from young plants (with the same sets of primers) was used, a portion of the CA-like gene was amplified and cloned. The portion cloned is the sequence of the 553 base pairs at the 3′ end of the gene. The 5′ end is still not found. The inframe STOP codon at the that is normally present at the 5′ end is not present in the sequence.

To look for what is upstream of the known sequence several techniques were used. They included Asymmetric PCR which uses one specific primer (antiparallel to the 5' primer of the CA-like sequence) and a non-specific primer at the 5' end. The second technique was to make a circle of DNA by intramolecular ligation. This circle would contain the known CA-like sequence and the unknown upstream region. The third technique was to amplify the upstream region by a primer antiparallel to the 5' primer of the CA-like sequence and another primer approximately 3600 base

pairs upstream. So far all these efforts were without good results. To shed light on what the upstream sequence might be we are also trying to align our sequence to others on the BCM search launcher (3). By creating overlaps of Expressed Sequence Tags (EST's) we should be able to work our way up to the beginning of our gene, maybe even up to the promoter sequences.

CAs in *Arabidopsis* can be "instructed" to enter the chlorolplast after they are synthesized (8). Conversely they can sometimes lack the chloroplast targeting sequence, serine and threonine rich sequence with few acidic amino acids (9.) In this case they function in the cytoplasm (10.) When we find the upstream sequence we will be able to tell where this particular β CA is located.

The CA-like sequence proposed by the ESSA group

ttg cttgtgcaga ctctagagtt tgtccttctg ctgtcctggg
attccaaccg ggtgacgcat tcactgttcg taacattgca aatttagtac ctccatatga
ggtataacat ttaatgcaga ctaccatttt ggttttatgg cctggttcac attttgatat
atttgcattt gcctgttgcg tgaagtctgg acctactgaa accaaagctg ctctagagtt
ctctgtgaat actcttaatg taagcatcta tcttatctct tatagtgttt cacgttttaa
atagagcatt tttcacagtt tagactgttc tgtttcttga aattagcttt ctacagagtt
tgtcttttca cttaactcaa atctctaggc aacataaaca gcatagacac aatgtgtaga
tacattcatt tataattctt gtctcattga taaacttgta ggtggaaaac atcttagtca
ttggtcatag ccggtgtgga ggaattcaag ctttaatgaa aatggaagac gaaggagatt
ccaggttttt accagttctt gattcaatct atcgaaaaaa tccattaatt gttgttgaac
tttgcccata ttgcatttcc ctggtaacag aagtttcata cacaactggg tagttgtggg
aaagaaggca aaggaaagca caaaagctgt tgcttcaaac ctccattttg atcatcagtg
ccaacattgt gaaaaggtga tgtcttttgg caaatcagcg gtctcttata tgaatcagtg
aacttatatt caatataaaa cctgcaggca tcgataaatc attcattaga aaggctgctt
gggtacccgt ggatagaaga gaaagtgcgg caaggttcac tgtctctcca tggtggatac
tataattttg ttgattgtac gttcgagaaa tggacagtgg attatgcagc aagcagaggt
aagaagaagg aaggcagtgg aatcgctgtt aaagaccggt cagtttggtc ttgac

Translation of the CA-like sequence

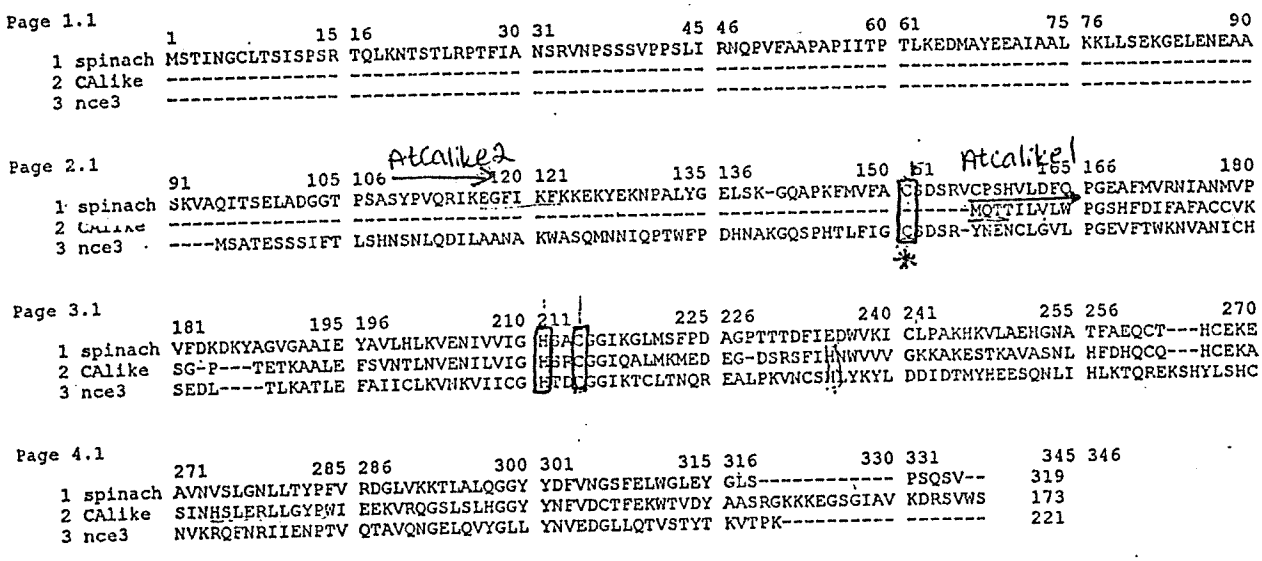
MQTTILVLWPGSHFDIFAFACCVKSGPTETKAALEFSVNTLNVE

NILVIGHSRCGGIQALMKMEDEGDSRSFIHNWVVVGKKAKESTKAVASNLHFDHQCQH

CEKASINHSLERLLGYPWIEEKVRQGSLSLHGGYYNFVDCTFEKWTVDYAASRGKKKE

GSGIAVKDRSVWS

Alignment with other Carbonic Anhydrases



Circled residues are necessary for CA activity

C* = cysteine residue in question.

Arrows represent the locations of the primers.

3.Materials and Methods

3.1 cDNA Library Amplification

A cDNA library (9-30-98) was first amplified by Polymerase Chain Reaction (PCR) with non-specific primers sp6 and T7. Then specific primers (AtCA-like-1, and -3; and AtCA-like-2, and -3)were used to amplify the desired sequence. The primers were diluted to 40 μM concentration. The enzyme used for amplification was Vent-polymerase. The products of the PCR reactions were tested by gel electrophoresis on 1.2% agarose gel (Figure 4.1a.) The marker used on the gel was lambda-HdIII. The products were purified with 50 mL phenol-chloroform premixed with isoamyl alcohol, they were precipitated with 3M NaOAc and washed with 100% and 70% ethanol. Poly-A-tails were added to the samples to prepare them for ligation into the vector (pGEM-easy.) The enzyme used for the A-tailing was Taq-polymerase. The enzyme used for ligation into the vector was T4 DNA ligase (3u / μL.) Then the vector and insert were transformed into competent DH5α *E. coli* cells by electroporation. The transformed cells were grown on two Ampicillin plates treated with X-gal. Five white colonies from each plate were then used to innoculate

growth medium tubes treated with 3µl ampicillin. The tubes were kept at 37 degrees Celsius overnight. The cultures then underwent alkline lysis, phenol-chloroform (isoamyl-alcohol) extraction, ethanol precipitation. The DNA samples were then raised in 50 µL of TE. To validate that the samples are correct the inserts were cut out of the vector using EcoRI. The samples were run on electrophoresis gel (Figure 4.1b.) Samples 1.4 and 1.5 were selected for preparation for sequencing. This preparation involved regrowing the overnight cultures, alkaline lysis, treatment with RNase, 100% ethanol wash, treatment with 13% PEG (and 4M NaCl), 70% ethanol wash, and raising in approximately 10 µL water to make the sample 1 mg / mL.

3.2 RNA Library Amplification

The next step was to try to get the sequence from an RNA library. The method used was Reverse Transcriptase PCR (RTPCR.) The enzyme MMLV-Reverse Transcriptase was used to make a DNA copy of the RNA strand. Then Taq Plus Precision DNA Polymerase was used to amplify the DNA strand. The primers were the

same as for the cDNA library. The products were run on 1.2% electrophoresis gel (Figure 4.2a.)

The products went through purification, precipitation, A-tailing, ligation into vector steps as described above. They were transferred into competent DH5α *E. coli* cells by electroporation. One white colony was picked from plate # 1, and 5 white colonies were picked from plate # 2. The colonies were used to inoculate overnight colonies as described in Section 3.1. Similarly, the cultures underwent alkaline lysis, phenol-chloroform (isoamyl-alcohol) extraction, ethanol precipitation, and raising in 50 μl of TE.
Then they were run on the agarose gel (Figure 4.2b.) To prove that the products were correct the DNA samples were cut with restriction enzymes XbaI, NcoI, HindIII and PstI, in addition to EcoRI (Figure 4.2c.)

Samples 2.1 and 2.2 were selected for preparation for sequencing as described in the above section.

3.3 Finding the 5′ end and the promoter sequence

After finding the sequence in isolated mRNA a new cDNA library was used to try to amplify the sequence upstream the CA-like gene. The goal was to find the 5′ end of the gene and the promoter and targeting sequences for the CA-like gene. To make sure that the CA-like sequence is present in the library PCR with pimers AtCA-like-2, and -3 and gel electrophoresis was carried out.

Three techniques were used to try to amplify the upstream sequence. One was Asymmetric PCR which uses one specific primer AtCA-like-4. This primer was antiparallel to the 5' primer of the CA-like sequence. The other primer was the non-specific primer T7 at the 5' end. The reaction was carried out at a temperature above the melting point (Tm) of the non-specific primer initially, so that the specific primer gets a "headstart" amplifying only the correct sequence.

The second technique was to make circle of DNA by intramolecular ligation. This circle would contain the known CA-like sequence and the unknown upstream region. In one of the cases the library amplified by sp6 and T7 was digested with SpeI to make "sticky ends"

for intramolecular ligation. The "circle-PCR" was also carried out without SpeI digestion.

The third technique was to amplify the upstream region by a primer antiparallel to the 5' primer of the CA-like sequence and another primer approximately 3600 base pairs upstream. For this a primer was designed at 3600 bases upstream the 5' end of the known sequence.

4. Results and Discussion

4.1 cDNA Amplification

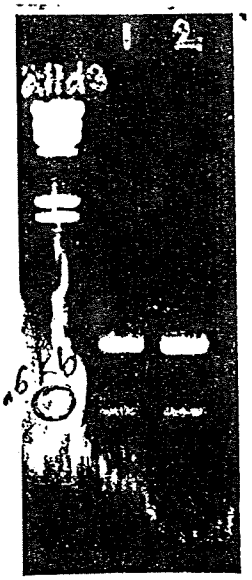


Figure 4.1a

The PCR reaction using primers AtCA-like-1 and AtCA-like-3, and AtCA-like-2 and AtCA-like-3 gave surprising results. The picture of the electrophoresis gel shows two bands flanking the 0.6 kb marker. The expected result was a single band higher than the 0.6 kb marker for primers AtCA-like-1, and -3. We expected to see a single band below the 0.6 kb marker for primers AtCA-like-2, and -3. The results were the same when the experiment was repeated.

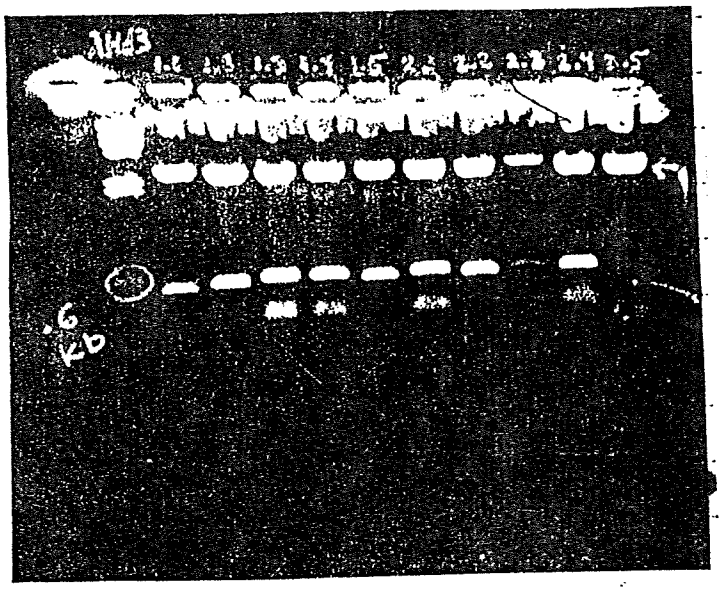


Figure 4.1b

After the DNA was reisolated from E. coli cells and digested with EcoRI the product was ran on the electrophoresis gel. The lanes indicate the 10 miniprep tubes (5 for first set of primers and 5 for the second set of primers.) The picture shows that some bands are single bands, what we expected in the beginning and some are double bands. The DNA samples in lanes 1.4 and 1.5 were prepared for sequencing. These samples were chosen because they represent the two types of results the gel showed.

Sequencing results of the segment amplified from the cDNA library:

The sequencing results indicated no homology to the proposed CA-like gene sequence. This means that the results are due to false priming. The reason we were not able to find the desired sequence may be that this protein is rare. An ideal cDNA library contains copies of every single mRNA in the cell. However in reality not all mRNA's are copied into cDNA or not in sufficient abundance.

4.2 RNA Library Amplification

Figure 4.2a

This picture shows the products of the RT-PCR. The bands should correspond to those in Figure 4.1a because the primers were the same. Lane #2 shows the same pattern but lane #1 shows a single band. Again, we expected to see a one band pattern. Lane C is the control lane.

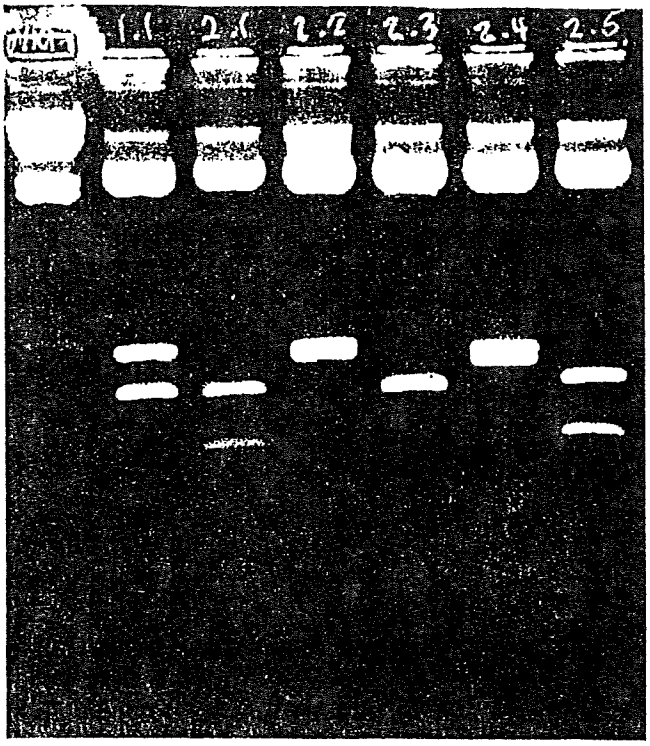


Figure 4.2b

This gel shows the digestion of the "miniprep" products with EcoRI. If there is only one band that suggests that there is no EcoRI site in the DNA sequence because all the restriction enzyme can do is cut the sequence out of the vector. (Once again EcoRI sites flank the insert in the vector.) If there are two bands that mean that there is a EcoRI site in our sequence. According to the restriction enzyme map created by the BCM Search Launcher there is an EcoRI site in the AtCA-like sequence. Therefor samples 2.1 and 2.5 were selected for further experiments.

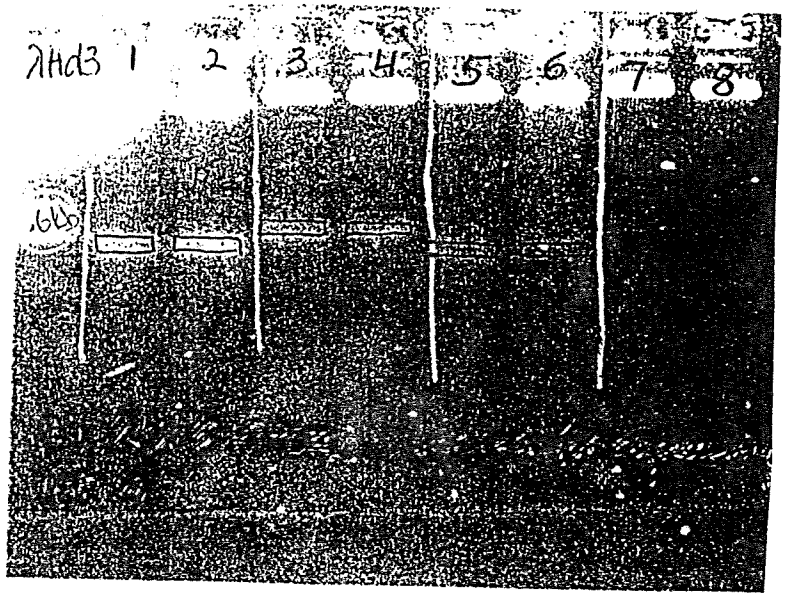


Figure 4.2c

This gel shows the products after digestion with several restriction enzymes. Lanes #1 and #2 show EcoRI digestion, lanes # 3 and #4 show NcoI digestion, Lanes #5 and #6 show HindIII and PstI digestion, and lanes #7 and #8 show digestion with XbaI. According to the restriction enzyme map (page 23) the following size fragments were expected. For EcoRI 200 and 375 bases, for HindIII and PstI 400 bases, for NcoI 450 bases, for XbaI 125 bases.

Sequencing results from isolated mRNA:

GTCAAGACCAAACTGACCGGTCTTTAACAGCGATTCCACTGCCTTCCTTCTTCTTACCTCTGCT
TGCTGCATAATCCACTGTCCATTTCTCGAACGTACAATCAACAAAATTATAGTATCCACCATGG
AGAGACAGTGAACCTTGCCGCACTTTCTCTTCTATCCACGGGTACCCAAGCAGCCTTTCTAATG
AATGATTTATCGATGCCTTTTCACAATGTTGGCACTGATGATCAAAATGGAGGTTTGAAGCAA
CAGCTTTTGTGCTTTCCTTTGCCTTCTTTCCCACAACTACCCAGTTGTGTATGAAACTTCTGGAA
TCTCCTTCGTCTTCCATTTTCATTAAAGCTTGAATTCCTCCACACCGGCTATGACCAATGACTA
AGATGTTTTCCACATTAAGAGTATTCACAGAGAACTCTAGAGCAGCTTTGGTTCCAGTAGGTC
CAGACTCATATGGAGGTACTAAATTTGCAATGTTACGAACAGTGAATGCGTCACCCGGTTGGA
ATCCCAGGACAGCAGAAGGACAAACTCTAGAGTCTGCCAAGCAA

Reverse Complement:
TTGCTTGGCAGACTCTAGAGTTTGTCCTTCTGCTGTCCTGGGATTCCAAC
CGGGTGACGCATTCACTGTTCGTAACATTGCAAATTTAGTACCTCCATAT
GAGTCTGGACCTACTGGAACCAAAGCTGCTCTAGAGTTCTCTGTGAATAC
TCTTAATGTGGAAAACATCTTAGTCATTGGTCATAGCCGGTGTGGAGGAA
TTCAAGCTTTAATGAAAATGGAAGACGAAGGAGATTCCAGAAGTTTCATA
CACAACTGGGTAGTTGTGGGAAAGAAGGCAAAGGAAAGCACAAAAGCTGT
TGCTTCAAACCTCCATTTTGATCATCAGTGCCAACATTGTGAAAAGGCAT
CGATAAATCATTCATTAGAAAGGCTGCTTGGGTACCCGTGGATAGAAGAG
AAAGTGCGGCAAGGTTCACTGTCTCTCCATGGTGGATACTATAATTTTGT
TGATTGTACGTTCGAGAAATGGACAGTGGATTATGCAGCAAGCAGAGGTA
AGAAGAAGGAAGGCAGTGGAATCGCTGTTAAAGACCGGTCAGTTTGGTCT
TGA

Alignment with the CA-like sequence proposed by the ESSA group

CA-like: ttg cttgtgcaga ctctagagtt tgtccttctg ctgtcctggg
Sequencing Result: ttg cttg gcaga ctctagagtt tgtccttctg ctgtcctggg

attccaaccg ggtgacgcat tcactgttcg taacattgca aatttagtac ctccatatga
attccaaccg ggtgacgcat tcactgttcg taacattgca aatttagtac ctccatatga

ggtataacat ttaatgcaga ctaccatttt ggttttatgg cctggttcac attttgatat
g ----------INTRON #1----------

atttgcattt gcctgttgcg tgaagtctgg acctactgaa accaaagctg ctctagagtt
----------tctgg acctactgaa accaaagctg ctctagagtt

ctctgtgaat actcttaatg taagcatcta tcttatctct tatagtgttt cacgttttaa
ctctgtgaat actcttaat----------

atagagcatt tttcacagtt tagactgttc tgtttcttga aattagcttt ctacagagtt
----------INTRON # 2----------

tgtcttttca cttaactcaa atctctaggc aacataaaca gcatagacac aatgtgtaga
----

tacattcatt tataattctt gtctcattga taaacttgta ggtggaaaac atcttagtca
----------tggaaaac atcttagtca

ttggtcatag ccggtgtgga ggaattcaag ctttaatgaa aatggaagac gaaggagatt
ttggtcatag ccggtgtgga ggaattcaag ctttaatgaa aatggaagac gaaggagatt

ccaggttttt accagttctt gattcaatct atcgaaaaaa tccattaatt gttgttgaac
ccag----------INTRON # 3----------

tttgcccata ttgcatttcc ctggtaacag aagtttcata cacaactggg tagttgtggg
----------aagtttcata cacaactggg tagttgtggg

aaagaaggca aaggaaagca caaaagctgt tgcttcaaac ctccattttg atcatcagtg
aaagaaggca aaggaaagca caaaagctgt tgcttcaaac ctccattttg atcatcagtg

ccaacattgt gaaaaggtga tgtcttttgg caaatcagcg gtctcttata tgaatcagtg
ccaacattgt gaaaag----------INTRON # 4 ----------

aacttatatt caatataaaa cctgcaggca tcgataaatc attcattaga aaggctgctt
----------gca tcgataaatc attcattaca aaccgtgctt

gggtacccgt ggatagaaga gaaagtgcgg caaggttcac tgtctctcca tggtggatac
gggtacccgt ggatagaaga gaaagtgagg caaggttcac tgtctctcca tggtggatac

tataattttg ttgattgtac gttcgagaaa tggacagtgg attatgcagc aagcagaggt
tataattttg ttgattgtac gttcgagaaa tggacagtgg attatgcagc aagcagaggt

aagaagaagg aaggcagtgg aatcgctgtt aaagaccggt cagtttggtc ttgac
aagaagaagg aaggcagtgg aatcgctgtt aaagaccggt cagtttggtc ttgac

The alignment shows the position of the introns. The sequence was amplified from isolted mRNA therefor it does not contain introns (shown in red.) The proposed "CA-like sequence" does contain introns (shown in black.) The position of the introns is shown in blue.

Restriction Map:

BsaMI
XbaI BsmI
ttgcttggcagactctagagtttgtccttctgctgtcctgggattccaaccgggtgacgcattcactgttcgtaa base pairs
aacgaaccgtctgagatctcaaacaggaagacgacaggaccctaaggttggcccactgcgtaagtgacaagcatt 1 to 75
Mva1269I

ApoI
BsrDI FauNDI XbaI
cattgcaaatttagtacctccatatgagtctggacctactggaaccaaagctgctctagagttctctgtgaatac base pairs
gtaacgtttaaatcatggaggtatactcagacctggatgaccttggtttcgacgagatctcaagagacacttatg 76 to 150
AcsI NdeI

Bse118I ApoI
BssAI AcsI HindIII
tcttaatgtggaaaacatcttagtcattggtcatagccggtgtggaggaattcaagctttaatgaaaatggaaga base pairs
agaattacacctttgtagaatcagtaaccagtatcggccacacctccttaagttcgaaattacttttaccttct 151 to 225
BsrFI EcoRI
Cfr10I

BbsI
BpuAI
cgaaggagattccagaagtttcatacacaactgggtagttgtgggaaagaaggcaaaggaaagcacaaaagctgt base pairs
gcttcctctaaggtcttcaaagtatgtgttgacccatcaacacccttttcttccgtttcctttcgtgttttcgaca 226 to 300
Bbv16II
BpiI

BspXI ClaI
Ksp22I Bsp106I
FbaI BanIII
tgcttcaaacctccattttgatcatcagtgccaacattgtgaaaaggcatcgataaatcattcattagaaaggct base pairs
acgaagtttggaggtaaaactagtagtcacggttgtaacacttttccgtagctatttagtaagtaatctttccga 301 to 375
BclI BspDI BseCI
Bsa29I
BscI Bsu15I

BshNI BstDSI NcoI Bsp19I
Asp718I EarI StyI DsaI
Eco64I DsaI Eam1104I Eco130I
gcttgggtacccgtggatagaagagaaagtgcggcaaggttcactgtctctccatggtggatactataattttgt base pairs
cgaacccatgggcacctatcttctctttcacgccgttccaagtgacagagaggtaccacctatgatattaaaaca 376 to 450
BanI KpnI Ksp632I ErhI BstDSI
Acc65I BssT1I
AccB1I EcoT14I

BcgI
tgattgtacgttcgagaaatggacagtggattatgcagcaagcagaggtaagaagaaggaaggcagtggaatcgc base pairs
actaacatgcaagctctttacctgtcacctaatacgtcgttcgtctccattcttcttccttccgtcaccttagcg 451 to 525

BsrFI
BssAI
AgeI Cfr10I
tgttaaagaccggtcagtttggtcttgac base pairs
acaatttctggccagtcaaaccagaactg 526 to 554
PinAI
BsaWI
Bse118I

Created with the BCM Search Launcher Web Cutter (11.)

Translation:

CLADSRVCPSAVLGFQPGDAFTVRNIANLVPPYESGPTGTKAALEFSVNTLNVENILVIGHSRCGGI
QALMKMEDEGDSRSFIHNWVVVGKKAKESTKAVASNLHFDHQCQHCEKASINHSLERLLGYPWI
EEKVRQGSLSLHGGYYNFVDCTFEKWTVDYAASRGKKKEGSGIAVKDRSVWS

The translation indicates the presence of the cysteine residue which is necessary for activity. (Shown in green)

5.Conclusion

In this research project a significant portion of a β Carbonic Anhydrase was amplified, cloned and sequenced from isolated mRNA from *Arabidopsis thaliana*. This β CA is located on chromosome #4. The sequence showed high homology to a proposed sequence for a "CA-like" gene published by the ESSA group. An important difference between the two sequences was that the translation of this β CA sequence shows the presence of a cysteine residue near the 5' end of the gene. This specific cysteine residue had been established to be indispensable for CA activity by Dr. Bartlett's laboratory. Another important piece of information we were able to gather from the sequencing data was the location of the splice sites in this gene.

We experienced many difficulties while we were trying to amplify this sequence. As a matter of fact we were never able to amplify it and clone it from a cDNA library. One of the main reasons that made our work challenging was the fact that this protein is rare in *Arabidopsis*. We never worked with genomic DNA. We only use a cDNA library and isolated mRNA, which both contain expressed sequences, therefor the level of

expression was a significant limiting factor in our chances of finding the desired sequence.

So far we know the sequence of the 554 bases at the 3′ end of this β CA in *Arabidopsis*. Time constraints did not allow us to finish looking for the 5′ end of the gene and its promoter sequence. However, I will continue working with Dr. Bartlett to solve the questions we still have. Once we find where the open reading frame begins we will look for the promoter and targeting sequences. With this information we will be able to use a reporting gene (Gus) in transgenic plants and we will be able to see the patterns of expression of this β CA.

6.Bibliography

(1) Bracey MH, Christiansen J, Tovar P, Cramer SP, Bartlett SG. (1994)*Biochemistry* **33:**13126-13131.

(2) Xi, Xioachun. Master of Science Thesis. Louisiana State University December 1999.

(3) http://www.hgsc.bcm.tmc.edu/SearchLauncher/

(4) Hewett-Emmet D, and Tashian RE. (1996) *Molec. Phylog. Evol.* **5**:50-77

(5)Sultemeyer D, Scmidt C, and Fork HP. (1993) *Physiologia Planatarum* **88**:179-190.

(6) Fawcett TW, Browse JA, Volokita M, Bartltett SG. *J. Biol Chem*, 1990 Apr 5

(7)http://www.arabidopsis.org

(8)Raines CA, PR, Holder C, Lloyd JC. *Plant Mol Biol* **20:**1143-1148.

(9) Johansson I, Forsman C. (1992) *FEBS Letters* **314**:232-236

(10)Fett JP, Coleman J. (1994) *Plant Phys* **105**:707-713

(11)http://dot.imgen.bcm.tmc.edu:9331/cgi-bin/seq-util/seq-util.pl